Hierarchical models for repeated measurements

Rien van der Leeden *

Abstract

Hierarchically structured data are common in many areas of scientific research. Such data are characterized by nested membership relations among the units of observation. Multilevel analysis is a class of methods that explicitly takes the hierarchical structure into account. Repeated measures data can be considered as having a hierarchical structure as well: measurements are nested within, for instance, individuals. In this paper, an overview is given of the multilevel analysis approach to repeated measures data. A simple application to growth curves is provided as an illustration. It is argued that multilevel analysis of repeated measures data is a powerful and attractive approach for several reasons, such as flexibility, and the emphasis on individual development.

Key words and phrases: repeated measures, growth curve analysis, longitudinal data, multilevel analysis, hierarchical linear model, hierarchical data.

*Department of Psychometrics and Research Methodology, Faculty of Social Sciences, University of Leiden, P.O. Box 9555, 2300 RB Leiden, The Netherlands.
Telephone: +31 71 5273763, Telefax: +31 71 5273619,
e-mail: vanderleeden@rulfsw.leidenuniv.nl

1 Introduction

Hierarchically structured data are frequently encountered in many areas of scientific research. Such data are characterized by so-called "nested" membership relations among the units of observation. For instance, in social and behavioural science research employees are nested within departments, or sportsmen are nested within teams. In biostatistics, birds are nested within breeding areas, or litters of offspring are nested within animals, and so on. Classic examples of hierarchical data are found in educational research: students are nested within classes, which are nested within schools. Many other examples can be imagined.

Multilevel analysis comprises a class of methods employing hierarchical linear regression models. Such models explicitly take into account the hierarchical structure of the data. Over the past 15 years, much progress has been made in the development of multilevel analysis. Originating from educational research (see, e. g., Burstein, Linn, & Capell, 1978; Tate & Wongbundit, 1983; De Leeuw & Kreft, 1986; Goldstein, 1987; Raudenbush, 1988; Bock, 1989), the technique is increasingly being used by now in numerous research settings. Comprehensive textbooks on theory and application of multilevel models include Goldstein (1987, 1995), Bryk and Raudenbush (1992) and Longford (1993a).

Traditional "single level" models fail when data are hierarchically structured, because the assumption of independence is violated. The nested structure causes so-called "intraclass dependency" among the observations within units at the higher level of the hierarchy. Multilevel analysis provides researchers with a very flexible and powerful set of tools to handle hierarchical data, with respect to both model formulation and hypothesis testing. From a statistical point of view, the application of multilevel models proves to be more precise (Goldstein, 1987; Bryk & Raudenbush, 1992). Moreover, it is argued that these models are conceptually enriching (Raudenbush, 1988).

In multilevel analysis, individuals are typically considered the first level units. Some kind of grouping defines the second level units. Similarly, more levels may be specified. Commonly, there are variables describing the individuals as well as describing the higher level units. A straightforward and most interesting generalization of multilevel analysis is in repeated measures data or growth curve analysis. Data from individuals that are measured at a number of consecutive points in time can be understood as having a two-level structure: measurements are nested within individuals. In this paper an overview is given of this approach to repeated measures analysis. New theory is not being developed, but known sources of information are combined to argue that the

approach is fruitful and attractive.

2 A two-level model

A fairly general two-level model for data obtained from N individuals, nested within J groups, each containing N_j individuals, is specified as follows. For each group j ($j = 1, \ldots, J$), the Level-1 or within-group model is

$$y_j = X_j\beta_j + e_j, \tag{1}$$

where y_j is a vector containing values on an outcome variable, X_j is a matrix of fixed, explanatory variables, β_j is a vector of regression coefficients and e_j is a vector of random error terms. At the second level, the Level-1 models are linked together by the Level-2 or between-group model, given by

$$\beta_j = W_j\gamma + u_j, \tag{2}$$

where W_j is a matrix of fixed, explanatory variables measured at the group level, γ is a vector containing fixed coefficients and u_j is a vector of random error terms. Usually, it is assumed that $e_j \sim N(0,\Sigma_{(1)})$ with $\Sigma_{(1)} = \sigma_e^2 I_{N_j}$, that $u_j \sim N(0,\Sigma_{(2)})$, and that the Level-1 random terms are distributed independently from the Level-2 random terms.

Equation 1 specifies a separate regression model for each group j with the same explanatory variables, but with a different set of regression coefficients. These coefficients can vary across Level-2 units, which is expressed in equation 2, where they are treated as random variables.

There are, of course, several ways to specify the Level-2 model. If W_j only consists of a vector of ones, the model specifies random variation of the coefficients across Level-2 units. Such models are referred to as *random coefficient models* (De Leeuw & Kreft, 1986; Prosser, Rasbash, & Goldstein, 1991). If W_j contains one or more Level-2 explanatory variables, part of the variation of the Level-1 coefficients may be explained. Further, the elements of β_j may be modelled differently, for instance intercepts can be random, whereas slopes are fixed.

Equations 1 and 2 give the illusion of a two-step procedure because it seems as if the Level-1 coefficients are regressed on the Level-2 explanatory variables within the Level-2 model. However, it is only conceptually attractive to specify the Level-1 and Level-2 models separately. The essential part of multilevel analysis is that we are considering the mixed linear model (cf. Harville, 1977), obtained by substituting equation 2 into 1,

written as

$$y_j = X_j W_j \gamma + X_j u_j + e_j. \tag{3}$$

Equation 3 contains a fixed part $X_j W_j \gamma$ and a random part $X_j u_j + e_j$. Note also that the fixed part contains (an) interaction term(s) between Level-1 and Level-2 explanatory variables. These terms are commonly called *cross-level* interaction terms. For some researchers cross-level interaction provides the major attraction to multilevel analysis.

The covariance matrix Σ_j of y_j, conditional to the fixed part, is expressed as

$$Var(y_j \mid X_j W_j \gamma) = Var(X_j u_j + e_j) = \Sigma_j = X_j \Sigma_{(2)} X_j' + \sigma_e^2 I_{N_j}. \tag{4}$$

Models for the complete data are obtained by stacking the J groups' models in 3 and 4. For 4, the resulting matrix Σ is block diagonal with blocks Σ_j. It follows that two sets of parameters have to be estimated: a set of fixed parameters and a set of random parameters, also called variance components. The fixed parameters are the elements of γ. The variance components are σ_e^2 at Level-1, and the elements of $\Sigma_{(2)}$ at Level-2.

Note that the within-group coefficients β_j seem to have been eliminated from the model. In fact, this is why multilevel analysis is usually compared favorably with separate (OLS) analyses within the different groups. The estimation of a large number of β's is replaced by the assumption of a distribution (usually joint normal) for these parameters over the groups, and the estimation of the parameters (γ and $\Sigma_{(2)}$) of this distribution. This makes the multilevel model far more parsimonious than separate models within the groups, and far more informative than one single model for the complete data ignoring grouping. Additional estimation of within-group coefficients, so-called "random Level-1 coefficients", is of course possible, but more or less apart from the mixed linear model described above.

3 A two-level model for repeated measures data

Repeated measures data, or growth data, can be described as measurements on one variable for the same (groups of) individuals on a number of consecutive points in time. It is very natural to extend the concept of hierarchy to this type of data, since repeated measurements are inherently dependent responses: they can be regarded as nested within individuals. In a two-level model it makes individuals the second level and measurements the first level. Models of this kind have already been around for some time. Key references include Laird and Ware (1982), Strenio, Weisberg, and Bryk (1983), Goldstein (1986a, 1986b, 1987, 1989), Bryk and Raudenbush (1987), Raudenbush (1989) and Bock (1989). Recent references include Bryk and Raudenbush (1992),

Hoeksma and Koomen (1992), Plewis (1994), Goldstein (1995), Rogosa and Saner (1995) and Snijders (1996).

Repeated measurements are almost always obtained for the assessment of change. Often, this concerns a kind of growth or development, for instance physical growth or learning achievement. Because such data reflect a developmental process as a function of time, a suitable method to model them is describing the expected values of the observations as functions of time. Polynomial functions comprise a class of functions often used for this purpose.

In a two-level model for repeated measurements, the Level-1 models specify unique (polynomial) growth curves, or growth trajectories, for each individual (Level-2 unit). At the second level the individual growth parameters are treated as random variables. The Level-2 model may simply model the growth parameters as an average over all individuals plus a person specific deviation. More elaborate Level-2 models include covariates as explanatory variables to account for between-subject variation in growth parameters. As such, systematic variation in growth trajectories can be studied, for instance, as a function of background variables and/or experimental treatments.

Usually, the model yields the estimated average growth trajectory over all individuals (possibly conditional on person-level covariates) and a set of estimated variance components. At Level-2 these components refer to the (conditional) between-subject variation in growth trajectories, whereas at Level-1, the within-subject variation not accounted for by the growth trajectory is estimated. In addition, it is possible to estimate the individual growth parameters.

A translation of this approach is that, for instance, differences in growth rate between different groups of individuals will be revealed not only in different mean levels of growth rate. Moreover, these differences will be modelled as different distributions of growth parameters within each group.

Analogously to equations 1 and 2, a two-level model for repeated measurements is specified as follows. The Level-1 or within-subject model is

$$y_j = T\beta_j + e_j, \tag{5}$$

where the vector y_j contains the repeated measurements for subject j, $(j = 1, \ldots, J)$, T is a matrix of known, constant variables (e. g. age in years, number of months, etc.) and possible transformations of these variables, β_j is a vector of individual parameters specifying the shape of the growth curve for subject j and e_j is a vector of random error

terms. More specifically, matrix T can be defined as

$$T = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^p \\ 1 & a_2 & a_2^2 & \cdots & a_2^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_t & a_t^2 & \cdots & a_t^p \end{pmatrix}$$

Matrix T shows that the degree of the polynomial fitted for subject j is p, but maximally $t-1$ (if there are t measurement occasions). Note that in this formulation, it is not necessay for each individual to have t measurements. Alternatively, T may either contain a set of standard, orthogonal polynomial vectors (see, e. g., Lury, 1950), or one might consider an orthogonalization of the original T.

The Level-2 or between-subject model is equal to 2, that is, $\beta_j = W_j\gamma + u_j$, but interpreted differently. Now W_j is a between-*subject* design matrix which can take several forms. If $W_j = I$, only simple random variation of the individual growth parameters is modelled. More elaborate models arise if W_j contains dummy variables coding subgroups of subjects and/or explanatory variables that may account for growth parameter variability. Covariates may be fixed across occasions (e. g. gender) or varying across time. So-called "time-varying" covariates can be included in T.

Now the mixed linear model for repeated measurements (substituting 2 into 5, cf. 3) is

$$y_j = TW_j\gamma + Tu_j + e_j, \tag{6}$$

where equal assumptions hold as for 3. The covariance matrix Σ_j of y_j, conditional to the fixed part, is now expressed as

$$Var(y_j \mid TW_j\gamma) = Var(Tu_j + e_j) = \Sigma_j = T\Sigma_{(2)}T' + \sigma_e^2 I_{N_j}. \tag{7}$$

One of the assumptions is that $e_j \sim N(0, \Sigma_{(1)})$ with $\Sigma_{(1)} = \sigma_e^2 I_{N_j}$. Equation 7 shows that for the application to repeated measures, this leads to a relatively simple covariance structure at the person-level: error terms are equal and uncorrelated across points in time. According to Bryk and Raudenbush (1987), it is very common to assume this simple structure when the number of measurement occasions is limited. They argue that with short time series, the assumption is practical and unlikely to corrupt results.

In some cases, however, it may be most useful to model a more complex Level-1 covariance structure. When there are many time points per subject, one may wish to model some form of time dependent or autocorrelation structure (see, e. g., Ware, 1985; Goldstein, Healy, & Rasbash, 1994). Other instances include studies where the

measurements are taken close together in time, or situations in which the error terms are dependent upon measured variables, such as age or time (Goldstein, 1986b, 1995). Error variance may also be group or person specific (Strenio et al., 1983). A model that can handle a variety of error covariance structures is derived by extending equation 6 to

$$y_j = TW_j\gamma + Tu_j + Z_j e_j, \tag{8}$$

where Z_j is a matrix of explanatory variables of which the coefficients are random at Level-1. The error covariance structure of model 8 is given by

$$Var(Tu_j + Z_j e_j) = \Sigma_j = T\Sigma_{(2)}T' + Z_j\Sigma_{(1)}Z_j', \tag{9}$$

Equation 9 contains a Level-2 and a Level-1 part. The covariance matrix $\Sigma_{(1)}$ is now unrestricted. Time dependent or autocorrelation structures may be modelled by appropriate specification of Z_j. In practice, matrices W_j, T and Z_j will often contain one or more of the same variables. Model 8 is the most flexible version of the longitudinal mixed linear model, because (compared to model 6) (1) explanatory variables can have random coefficients at Level-1 as well, and (2) explanatory variables in the random part of the model not necessarily have been included in the fixed part.

The topic of *centering* concerns multilevel models in general. It means that one has to decide upon an appropriate location of the variables in a model. The choice of a particular location will determine the way in which the coefficients should be interpreted. One of the most important things to consider is the meaning of a score of zero on the predictor variable(s) (for instance, in a simple Level-1 model with one predictor this determines the interpretation of the intercept).

Generally, there are several possibilities for the location of the predictors, such as centering around the grand mean, or around the Level-2 means, or no centering at all (natural metric). For longitudinal applications, it may be useful to center the Level-1 predictors in such way that the intercept equals the expected outcome for subject j at a specific point in time, for instance, at the start of the training program. Comprehensive discussions on centering in multilevel analysis can be found in Bryk and Raudenbush (1992), Longford (1993a) and Kreft, De Leeuw and Aiken (1995).

Relations with MANOVA and GLMs

Models 6 and 8 are related to the more familiar MANOVA-based treatment of growth curve data in the following way. MANOVA approaches to repeated measurements have in common that for a group of J individuals the observations on the t different points in time are conceived as a set of dependent variables. Hence, models are of the form

$E[Y] = XM$, where Y and X are the complete response matrix and between-subject design matrix, respectively, and M is a matrix of coefficients. Wishart (1938) was one of the earliest to present a class of models in which MANOVA and polynomial curve fitting were connected. His approach was to find growth curves consisting of a set of least squares estimates of orthogonal polynomial coefficients. Represented in some defined experimental design, these curves could be subject to univariate or multivariate analysis of variance. MANOVA, considered superior to ANOVA for this purpose (cf. Visser, 1985), was combined with polynomial curve fitting in one single model by Potthoff and Roy (1964). Their model is of the form $E[Y] = XBT$, where Y and X are the response and design matrix, B is a parameter matrix and T is a matrix containing (orthogonal) polynomials (compare T in models 6 and 8). Potthoff and Roy presented their model as a generalized MANOVA model. This generalization is however limited to the mean structure, that is, the fixed part of the model. The structure of the covariances (the random part, compare Σ, obtained by stacking the J groups' Σ_j in equations 7 and 9) was in fact left unspecified. For most variations of the Potthoff-Roy model it holds that they mainly focus on the structure for the mean responses: mean curves for specified groups are estimated, and usually simple structures for Σ are employed. Besides, estimation of individual growth curves receives less attention as well.

For all J individuals, the longitudinal mixed linear models 6 and 8 can be written as $E[Y] = (W\Gamma)'T'$, showing the close similarity of the models' fixed part to the generalized MANOVA model. More recent generalizations of the MANOVA repeated measures model include more elaborated structures for Σ_j. Probably, the most comprehensive model is discussed by Jennrich and Schluchter (1986) (implemented as procedure `BMDP-5V` in the `BMDP` package). Their model is of the form $Y = (X\Gamma)' + UV + E$, where X contains within-subject (the polynomials in T), as well as between-subject (W_j) explanatory variables, and V may contain T. Writing model 6 as $Y = (W\Gamma)'T' + UT + E$ shows the close similarity of both models. The Jennrich-Schluchter model shows where MANOVA repeated measures models and random coefficient models meet, although it is formulated from within a MANOVA frame of reference (for a more detailed discussion see Van der Leeden, Vrijburg, & De Leeuw, 1996). Model 8 is slightly more general as far as the random part is concerned.

The longitudinal mixed linear models 6 and 8 could also be derived as special cases within the class of "generalized linear models" (GLMs) (see, e. g., McCullagh & Nelder, 1989). Zeger, Liang, and Albert (1988), for instance, study a mixed generalized linear model for the analysis of repeated measures data. GLMs are ususally fitted using the `GLIM` package (Francis, Green, & Payne, 1993) which has been designed as an interactive

"tool kit" for statistical modelling. As a result, longitudinal mixed generalized linear models should be developed (programmed) as GLIM macros.

4 Estimation and testing

Multilevel models can be fitted by several computer programs widely available today. Major packages include ML3 (Prosser et al., 1991), MLn (Rasbash, Yang, Woodhouse, & Goldstein, 1995), VARCL (Longford, 1993b) and HLM (Bryk, Raudenbush, & Congdon, 1996). Multilevel-type procedures are also available in SAS (PROC MIXED) and, as mentioned above, in BMDP (BMDP-5V) (for a comparative study of ML3, VARCL, HLM and BMDP-5V, see Kreft, De Leeuw, & Van der Leeden, 1994). They could also be developed within the GLIM package, but this is less straightforward.

Estimates of the fixed parameters and the variance components of the mixed linear models 6 or 8 (stacking the J groups' models) can be obtained by minimizing the deviance indicated by function $\triangle$, written as

$$\triangle(\gamma, \Sigma \mid y) = c + \log|\Sigma| + (y - TW\gamma)'\Sigma^{-1}(y - TW\gamma), \tag{10}$$

where Σ has structure 7 or 9.

Several procedures to minimize $\triangle$ have been developed and implemented in the available software. In one way or another, these procedures are all versions of full information maximum likelihood (FIML), or restricted maximum likelihood (REML) (if multivariate normality is assumed, the deviance equals twice the negative of the logarithm of the likelihood function for these models). REML methods optimize the likelihood function conditional on the fixed parameters, that is, with respect to the variance components only. It is common knowledge that FIML and REML estimators have several attractive properties, such as consistency and efficiency. A specific advantage of REML methods is that they provide unbiased estimators of the variance components (e. g., Searle, Casella, & McCulloch, 1992, Ch. 6). A drawback of both approaches is that generally the parameter estimates must be obtained iteratively.

ML procedures include the EM algorithm (cf. Dempster, Rubin, & Tsutakawa, 1981), implemented in HLM, and the method of Fisher scoring (Longford, 1987), implemented in VARCL. Goldstein (1986b) developed an iterative generalized least squares (IGLS) procedure, and a restricted version (RIGLS). Both methods are implemented in ML3 and MLn. If normality assumptions are met, IGLS and RIGLS are equivalent to FIML and REML, respectively. An alternative approach is Bayes estimation (Lindley & Smith,

1972; Strenio, 1981). For a detailed treatment of the various estimation procedures see Bryk and Raudenbush (1992) and Longford (1993a).

Additional to the estimation of fixed parameters and variance components, the random Level-1 coefficients, β_j, can be estimated by the method of shrinkage. In repeated measures applications, β_j is the set of growth parameters for subject j. The shrinkage estimator $\hat{\beta}_j^*$ is written as

$$\hat{\beta}_j^* = \hat{\Lambda}_j \hat{\beta}_j + (I - \hat{\Lambda}_j) W_j \hat{\gamma}, \tag{11}$$

where $\hat{\Lambda}_j$ is an estimated "reliability" matrix (cf. Bryk & Raudenbush, 1992, p. 43) and $\hat{\beta}_j$ is the OLS estimator of β_j based on the data for group/subject j. Shrinkage estimators combine the two available sources of information for estimation of β_j: the within-group/subject OLS estimates, and the overall sample estimates $W_j\hat{\gamma}$. The more reliable the OLS estimates, the more weight is put upon them, and vice versa. Hence, the unbiased but inefficient estimator $\hat{\beta}_j$ is "shrunk" towards the biased but stable estimator $W_j\hat{\gamma}$ (cf. Longford, 1993a, p. 17). The estimator $\hat{\beta}_j^*$ is also called an *empirical Bayes estimator* (Morris, 1983).

Single-parameter tests for fixed parameters and variance components can be derived as follows (for a comprehensive discussion of hypothesis tests see Bryk & Raudenbush, 1992).

Fixed parameters. The ususal null hypothesis for a fixed parameter is $H_0 : \gamma_g = 0$. This amounts to testing the hypothesis that a specific person-level predictor variable (in W_j) has no effect on a corresponding growth parameter. H_0 may be tested using a z-ratio defined as

$$z = \hat{\gamma}_g / (\hat{V}_{\gamma_g})^{1/2}, \tag{12}$$

where $\hat{\gamma}_g$ is the ML estimate of γ_g and $\hat{V}_{\gamma_g}$ is the estimated sampling variance of $\hat{\gamma}_g$. Bryk and Raudenbush, however, argue that in practice a t-ratio with degrees of freedom equal to $N - p - 1$ will often give more reliable results (N is the number of Level-2 units, in our case subjects, and p is the number of person-level predictors, except for the unit variable associated with the intercept) (see Bryk & Raudenbush, 1992, p. 50). This will be particularly important when the number of Level-2 units is small. Only the HLM program provides p-values based on the t, the other programs just give standard error estimates.

Variance components. A hypothesis concerning a variance component could be, for instance, the null hypothesis that there is no significant variation in the slope parameters across subjects. Such hypothesis can be written as $H_0 : \sigma_g^2 = 0$, where $\sigma_g^2 = Var(\beta_{gj})$. Generally, if Level-2 units contain enough Level-1 units to justify the computation of the

OLS estimates $\hat{\beta}_j$, that is here, if the number of measurements per subject is sufficiently large, the following statistic can be computed. Let $\hat{d}_{gj}$ denote the g'th diagonal element of $\hat{D}_j = \hat{\sigma}_e^2(T'T)^{-1}$. Then the statistic

$$\sum_j (\hat{\beta}_{gj} - \hat{\gamma}_{g0} - \hat{\gamma}_{g1} Z_{1j} - \cdots - \hat{\gamma}_{gp} Z_{pj})^2 / \hat{d}_{gj}, \tag{13}$$

will be approximately χ^2 distributed with degrees of freedom equal to $N - p - 1$ (see Bryk & Raudenbush, 1992, p. 55).

Sometimes the same hypothesis is tested using a z-ratio based on the estimated standard error of $\hat{\sigma}_g^2$. However, in most instances the normality approximation will be very bad and so these z-tests are not well-founded.

Generally, likelihood-ratio tests may be the most preferable to test hypotheses about the variance components. These tests are well known from standard literature. They are derived by computing the difference between deviance-values from "nested models" and using a χ^2 distribution as a reference (see e. g. Wilks, 1962, p. 55).

5 Illustration: rat growth

The multilevel model for repeated measures data is illustrated with a set of data on rat growth originally presented and analyzed by Box (1950) and afterwards by several others (Rao, 1965; Hills, 1974; Strenio et al., 1983). The data analyzed here are taken from Strenio, Weisberg and Bryk (1983). These authors used only a part of the original set, and because their purpose was to provide an illustrative analysis as well, they added some random error and constructed a covariate. The data consist of the weights of ten rats (WEIGHT), measured initially and after each of four consecutive weeks. The covariate is mother's weight (MW). Although simple, this example analysis serves to illustrate the multilevel approach to repeated measurements.

Results from the literature (mentioned above) suggest fitting a linear growth model. The variable WEEK has the values 0,1,2,3,4 and 5. The within-rat (Level-1) model can be written as

$$(WEIGHT)_{jt} = \beta_{0j} + \beta_{1j}(WEEK)_{jt} + e_{jt} \tag{14}$$

and the between-rat (Level-2) model is

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}(MW)_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(MW)_j + u_{1j}. \end{aligned} \tag{15}$$

Table 1: *Rat data, parameter estimates for linear growth model with covariate*

Fixed parameter	Estimate	se	t ratio
γ_{00}	12.940	36.300	0.356
γ_{10}	0.251	0.223	1.126
γ_{01}	2.967	11.850	0.250
γ_{11}	0.147	0.073	2.014
Variance component	**Estimate**	**se**	
σ_e^2	91.750	23.690	
σ_0^2	82.840	63.210	
σ_{01}	-19.270	19.120	
σ_1^2	5.514	6.975	
deviance	376.72		

The single-equation specification of the model is

$$\begin{aligned}(WEIGHT)_{jt} &= \gamma_{00} + \gamma_{01}(MW)_j + \gamma_{10}(WEEK)_{jt} \\ &\quad +\gamma_{11}(WEEK)_{jt}(MW)_j \\ &\quad +u_{0j} + u_{1j}(WEEK)_{jt} + e_{jt}. \end{aligned} \tag{16}$$

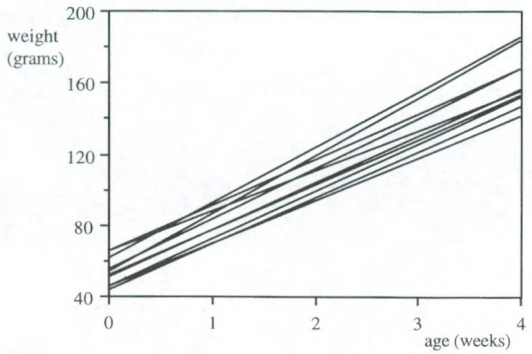
Parameter estimates for model 16 have been obtained with `ML3` using the `RIGLS` method. Results are given in Table 1.

Note that σ_e^2 is the Level-1 variance component, whereas σ_0^2, σ_{01} and σ_1^2 are the Level-2 variance components (the elements of $\Sigma_{(2)}$). Hence, σ_0^2 and σ_1^2 denote the variance of the intercepts (initial growth status) and the variance of the slopes (growth rate), respectively, that is, the variance not accounted for by `MW`; σ_{01} denotes the covariance between both growth parameters.

Because sample size is extremely small here, we may argue the usefulness of the sigificance tests. However, neglecting this problem for the sake of our illustrative purposes, the results can be interpreted in the following way.

Standard errors and t-ratio's in Table 1 indicate that except for the 'interaction' parameter γ_{11}, none of the fixed parameters is significant. Fitting a random coefficient growth model without the covariate `MW` (not reported here) shows a significant slope. Thus, correcting for mother's weight reduces the significance of this parameter, but keeps the linear trend (γ_{10}) unimpaired, though weeker. The fixed parameters further

Figure 1: *Rat data, predicted growth curves for 10 rats*



show that MW has a positive effect on the initial rat weight (γ_{01}), as well as on the growth rate (γ_{11}). This means that rats with a stout mother have a higher initial weight and gain weight faster.

It is difficult to give an interpretation of the variance component estimates themselves, since their value depends on the scaling of the variables. Nevertheless, comparing results of the random coefficient growth model without MW (not reported here) shows that the covariate has a strong effect on the growth trajectories: MW appears to explain a substantial amount of the Level-2 variance, that is, the variance of the initial weights and the growth rates. The negative covariance between initial status and growth rate is sometimes interpreted as that generally subjects (rats) that start at a low level tend to develop faster than others that start at a higher level.

Additionally, shrinkage estimates ($\hat{\beta}_j^*$) of the growth parameters were computed for each rat. These are used to plot the predicted growth curves in Figure 1.

6 The merits of longitudinal multilevel analysis

Multilevel analysis of repeated measures data has several attractive qualities which will be discussed below.

Attributes of growth. Every multilevel polynomial growth curve model, even the most simple, includes two important characteristics of growth or development. First, the (co)variances among the observations are a function of time. This can be seen from equation 7 and it is easily illustrated by considering the simple random coefficient linear growth model $Y_{jt} = \gamma_{00} + \gamma_{10}a_{jt} + u_{0j} + u_{1j}a_{jt} + e_{jt}$. Now let $r_{jt} = u_{0j} + u_{1j}a_{jt} + e_{jt}$, the random part of the model. Then, for any subject j, the (co)variances among the elements of y_j are given by

$$Var(r_{jt}) = \sigma_0^2 + 2\sigma_{01}a_{jt} + \sigma_1^2 a_{jt}^2 + \sigma_e^2. \tag{17}$$

Hence, the between-subject variation is a function of time (or age), which makes sense because subjects are expected to grow at different rates.

Second, the responses of a subject on different occasions are correlated, that is, every two observations for a single subject are correlated. This characteristic refers to the serial dependency in repeated measurements. It can be shown by expressing the covariance between the observations of subject j at occasions t and t' as

$$Cov(r_{jt}, r_{jt'}) = \sigma_0^2 + (a_{jt} + a_{jt'})\sigma_{01} + a_{jt}a_{jt'}\sigma_1^2. \tag{18}$$

Equation 18 shows that the covariance between every two responses depends on the spacing of the observations, on the relative size of the variances of both the intercepts and the slopes, and on the covariance between them.

Bryk and Raudenbush (1992) present the following topics to characterize the favorable properties of longitudinal multilevel analysis, and to compare it with the usual MANOVA approach.

Emphasis on individual growth. The multilevel model for repeated measurements explicitly takes into account the individual growth. The starting point of the model is the Level-1 model representing the individual growth trajectory. In common MANOVA repeated measures methods, individual variation in growth trajectories is not modelled directly, but derived from the interaction of repeated measures by individuals. One could argue that the hierarchical approach is therefore conceptually more suited for growth curve modelling (see, e. g., Willett, 1988).

Flexibility of approach. Generally, the longitudinal multilevel model can be considered more flexible than the MANOVA repeated measures model. Within the multilevel framework, it is possible to formulate a variety of growth models. Polynomial growth curves can be of any degree and there is maximal freedom to chose predictor variables in the person-level model for each growth parameter separately. Including time-varying

covariates (covariates that vary across occasions) in the Level-1 model is straightforward. Time variables do not necessarily need to be discrete but may also be continuous. Especially noteworthy is the fact that the number of observations per subject, as well as the spacing of these observations in time may vary. This means that missing data can be handled very well.

Modelling of the within-subject covariance structure. The longitudinal multilevel model allows for the modelling of the within-subject covariance structure (cf. 9). It is possible to include explanatory variables that (partially) account for these (co)variances and to model specific structures, such as autocorrelation structures. Using MANOVA repeated measures methods it is usually impossible to model the within-subject covariance structure. The Jennrich-Schluchter model implemented in `BMDP-5V`, was the first MANOVA repeated measures approach in which (specific) covariance structures could be modelled directly (Jennrich & Schluchter, 1986; Dixon, 1988).

Correspondence in results. Concerning data requirements (e. g. balancedness) and assumptions (e. g. about the within-subject covariance structure), conventional MANOVA repeated measures methods are more restrictive than longitudinal multilevel models. It can be shown, however, that when MANOVA requirements are satisfied, longitudinal multilevel models give the same estimates for fixed effects and the same t-ratio's as MANOVA repeated measures procedures.

Models with more than two levels. The hierarchical approach to repeated measures data permits the formulation not only of two-level models, but also of higher-level models. One could think of a third level, for instance, at which the subjects are clustered. In such models, the effect of "organization" on growth can be studied.

Acknowledgements

I am indebted to the referee for his helpful comments. Jord Kooijman is gratefully acknowledged for his assistance in preparing the example analyses.

References

Bock, R. D. (1989). *Multilevel analysis of educational data.* San Diego, CA: Academic Press.

Box, G. E. P. (1950). Problems in the analysis of growth and wear curves. *Biometrics*, *6*, 362–389.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierachical linear models: applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM. Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs* (updated for version 4, 2nd ed.). Chicago, IL: Scientific Software International, Inc.

Burstein, L., Linn, R., & Capell, F. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics, 3*, 347–383.

De Leeuw, J., & Kreft, I. G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*, 57–85.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association, 76*, 341–353.

Dixon, W. J. (1988). *BMDP Statistical software manual, vol. 2.* Berkely, CA: University of California Press.

Francis, B., Green, M., & Payne, C. (1993). *The GLIM System, Release 4 manual.* Oxford, GB: Clarendon Press.

Goldstein, H. (1986a). Efficient statistical modelling of longitudinal data. *Annals of Human Biology, 13*, 129–141.

Goldstein, H. (1986b). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43–56.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London, GB: Oxford University Press.

Goldstein, H. (1989). Models for multilevel response variables with an application to growth curves. In R. Bock (Ed.), *Multilevel analysis of educational data* (pp. 107–125). San Diego, CA: Academic Press.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London, GB: Edward Arnold.

Goldstein, H., Healy, M., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, *13*, 1643–1655.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*, 320–340.

Hills, M. (1974). *Statistics for comparative studies.* New York, NY: John Wiley.

Hoeksma, J., & Koomen, M. (1992). Multilevel models in developmental psychological research: rationales and applications. *Early Development and Parenting*, *1*, 157–167.

Jennrich, R., & Schluchter, M. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, *42*, 805–820.

Kreft, I. G. G., De Leeuw, J., & Aiken, S. L. (1995). The effects of different forms of centering in hierarchical lineair models. *Multivariate Behavioral Research*, *30*, 1–22.

Kreft, I. G. G., De Leeuw, J., & Van der Leeden, R. (1994). A review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *The American Statistician*, *48*, 324–335.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society*, *B34*, 1–41.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*, 817–827.

Longford, N. T. (1993a). *Random coefficient models.* Oxford, GB: Clarendon Press.

Longford, N. T. (1993b). *VARCL. Software for variance component analysis of data with nested random effects (maximum likelihood).* Groningen, The Netherlands: iec ProGAMMA.

Lury, D. B. de. (1950). *Values and integrals of the orthogonal polynomials up to n = 50.* Toronto: University of Toronto Press.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London, GB: Chapman & Hall.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association, 78*, 47–65.

Plewis, I. (1994). Longitudinal multilevel models. In A. Dale & R. Davies (Eds.), *Analyzing social and political change* (pp. 118–135). London: Sage.

Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*, 313–326.

Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3. Software for three-level analysis. user's guide for V. 2.* London, GB.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52*, 447–458.

Rasbash, J., Yang, M., Woodhouse, G., & Goldstein, H. (1995). *MLn: command reference guide.* London, GB.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics, 13*, 85–1160.

Raudenbush, S. W. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research, 13*, 721–740.

Rogosa, D. R., & Saner, H. (1995). Longitudinal data analysis examples with random coefficient models. *Journal of Educational and Behavioral Statistics, 20*, 149–170.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York, NY: John Wiley.

Snijders, T. A. B. (1996). Analysis of longitudinal data using the hierarchical linear model. *Quality and Quantity, 30*, 405-426.

Strenio, J. F. (1981). *Empirical Bayes estimation for a hierarchical linear model.* Unpublished doctoral dissertation, Harvard University.

Strenio, J. F., Weisberg, H. I., & Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics, 39*, 71–86.

Tate, R. L., & Wongbundit, Y. (1983). Random versus non-random coefficient models for multilevel analysis. *Journal of Educational Statistics, 8*, 103–120.

Van der Leeden, R., Vrijburg, K., & De Leeuw, J. (1996). A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: comparing hierarchical linear regression (ML3, HLM) and repeated measures design with structured covariance matrices (BMDP5V). *Computational Statistics and Data Analysis, 21*, 581–605.

Visser, R. A. (1985). *Analysis of longitudinal data in behavioural and social research.* Leiden, The Netherlands: DSWO Press.

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *Journal of the American Statistical Association, 80*, 95–101.

Wilks, S. (1962). *Mathematical statistics.* New York, NY: John Wiley.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Rithkopf (Ed.), *Review of research in education (1988-89)* (pp. 345–422). Washington, DC: American Educational Research Association.

Wishart, J. (1938). Growth rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika, 30*, 16–28.

Zeger, S. L., Liang, K., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics, 44*, 1049–1060.

Ontvangen: 08-05-1996
Geaccepteerd: 01-12-1996